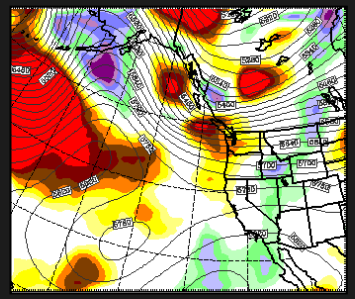
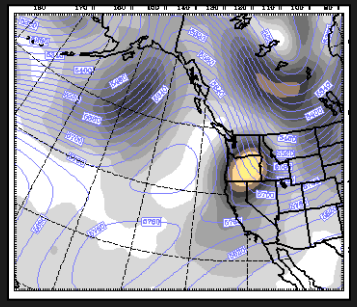
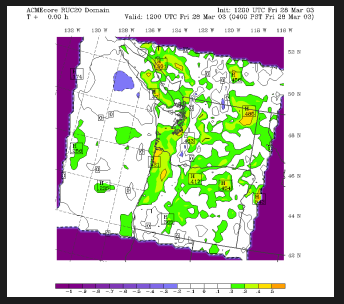
UWME and Probabilistic Spread-Skill Relationships

Eric P. Grimit

University of Washington







Acknowledgements:

Clifford F. Mass

F. Anthony Eckel

DoD MURI administered by ONR

A Consortium of Federal and Local Agencies

National Weather Service

Part I:

The University of Washington Mesoscale Ensemble (UWME)

A Little History

- The UW mesoscale ensemble (UWME) system was borne out of experience with a high-resolution (down to 4-km) local MM5 effort.
- Specifically, although high resolution produced better (sharper, high amplitude) structures in general, the forecasts verified only marginally better than lower resolution forecasts using traditional measures.
- UW research revealed large differences, and thus uncertainty, in the initializations of, and MM5 forecasts based on, major operational forecasting systems [McMurdie and Mass, WAF 2004].
- Subjective verification showed that approximately 12-km grid spacing was needed to capture major regional mesoscale features.
- Thus, it was natural to create a 12-km mesoscale ensemble system for the Northwest.

The UW SREF Approach

- Create an effective mesoscale SREF system capable of:
  - producing skillful forecast probability
  - providing researchers with a rich data set to answer some basic questions:
    - Does an ensemble need to be made up of equally likely solutions to be useful?
    - How much can be accomplished with easily acquired resources?
    - What is the relative importance of model inadequacy compared to initial uncertainty?

- The basic idea is that differences in the analyses of various operational centers are a valuable measure of IC uncertainty.
  - Analysis differences reflect different data inventories, assimilation schemes, and model physics/numerics.
  - Analysis differences can be large, often greater than observation errors themselves.
  - Designed a multi-analysis-based ensemble system to capitalize on these differences.
    - 5-member multi-analysis system, 2000-01 (Grimit and Mass, WAF 2002)
    - 7-members, 2001-02 (Mass et al., BAMS 2003)
    - 8-members, 2002-current (Eckel and Mass, accepted WAF)

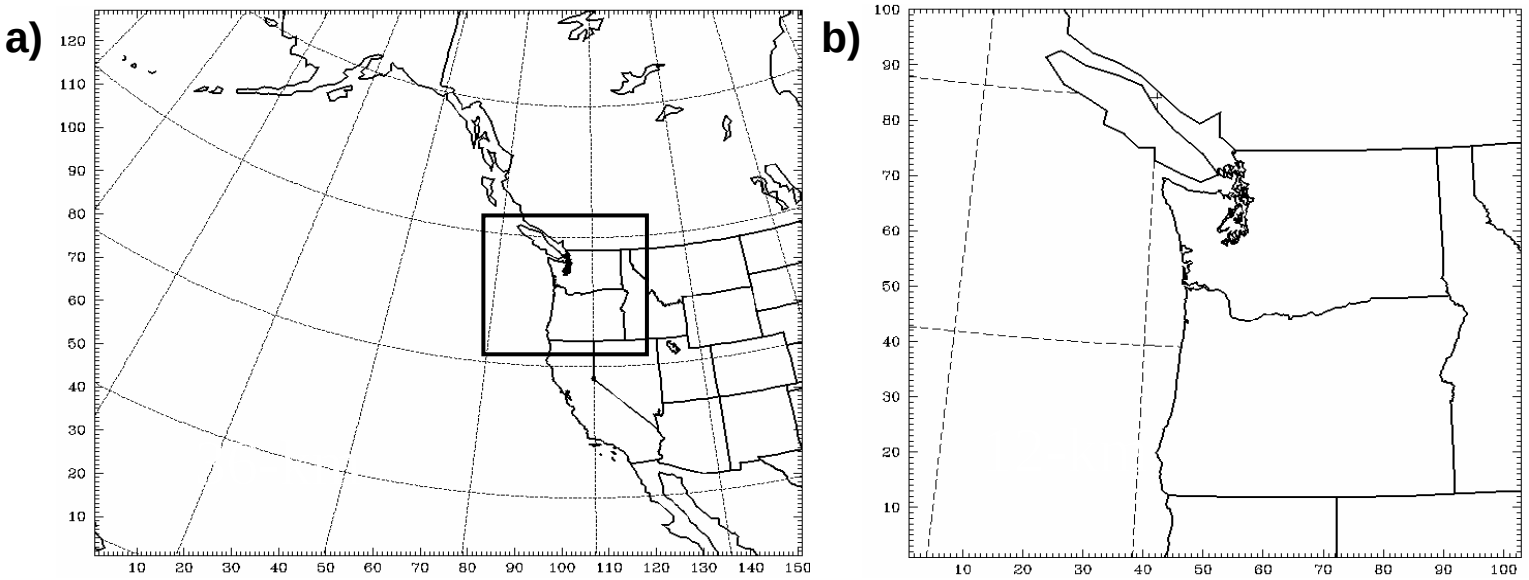
- Encouraged by contemporaneous work by David Richardson (ECMWF), Beth Ebert (BOM-Australia), and David Baumhefner (NCAR).
  - Multi-analysis ensembles competitive and appropriate in short-range
  - IC perturbations at synoptic scales yield predictability error growth at ALL scales

Multi-Analysis Sources

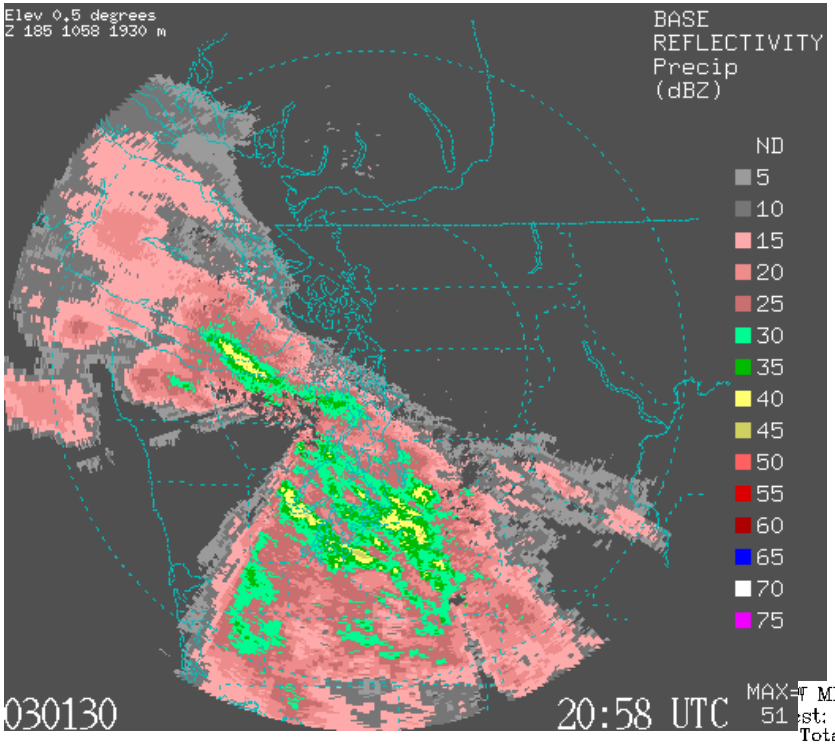
Abbreviation/Model/Source	Type	Resolution (~ @ 45 °N) Computational	Resolution (~ @ 45 °N) Distributed	Objective Analysis
avn, Global Forecast System (GFS), National Centers for Environmental Prediction	Spectral	T254 / L64 ~55 km	1.0° / L14 ~80 km	SSI 3D Var
cmcg, Global Environmental Multi-scale (GEM), Canadian Meteorological Centre	Finite Diff	0.9°×0.9°/L28 ~70 km	1.25° / L11 ~100 km	3D Var
eta, limited-area mesoscale model, National Centers for Environmental Prediction	Finite Diff.	32 km / L45	90 km / L37	SSI 3D Var
gasp, Global AnalysiS and Prediction model, Australian Bureau of Meteorology	Spectral	T239 / L29 ~60 km	1.0° / L11 ~80 km	3D Var
jma, Global Spectral Model (GSM), Japan Meteorological Agency	Spectral	T106 / L21 ~135 km	1.25° / L13 ~100 km	OI
ngps, Navy Operational Global Atmos. Pred. System, Fleet Numerical Meteorological & Oceanographic Cntr.	Spectral	T239 / L30 ~60 km	1.0° / L14 ~80 km	OI
tcwb, Global Forecast System, Taiwan Central Weather Bureau	Spectral	T79 / L18 ~180 km	1.0° / L11 ~80 km	OI
ukmo, Unified Model, United Kingdom Meteorological Office	Finite Diff.	5/6°×5/9°/L30 ~60 km	*same* / L12	3D Var

UW Mesoscale Ensemble System

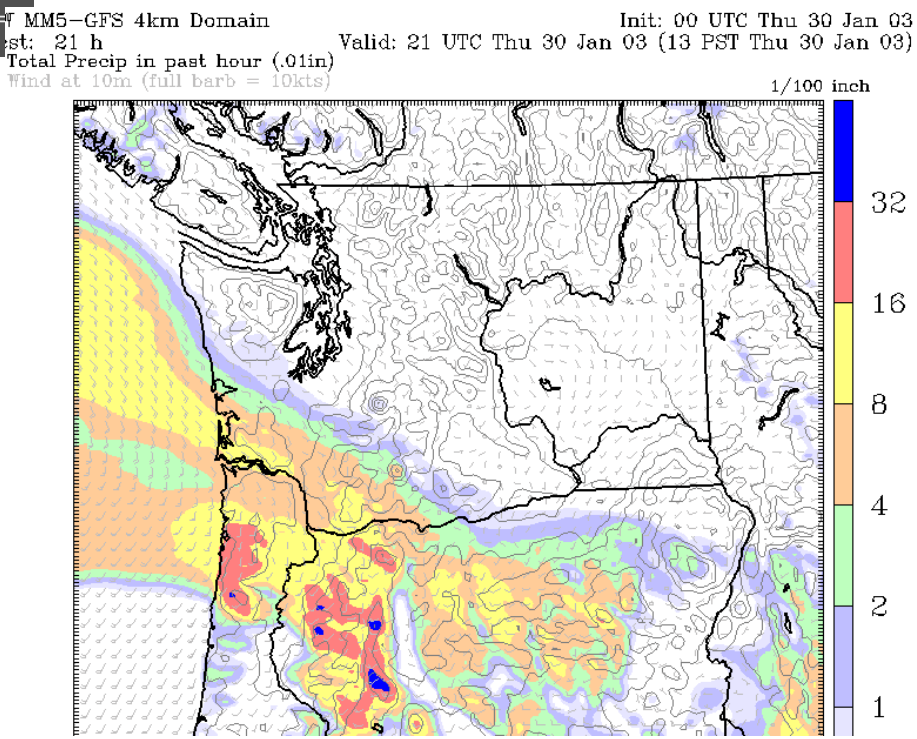
- Limited-area mesoscale modeling system (MM5)
- 2-day (48-hr) forecasts at 0000 UTC in real-time
  - Since January 2000
- Current 36-km and 12-km domains:



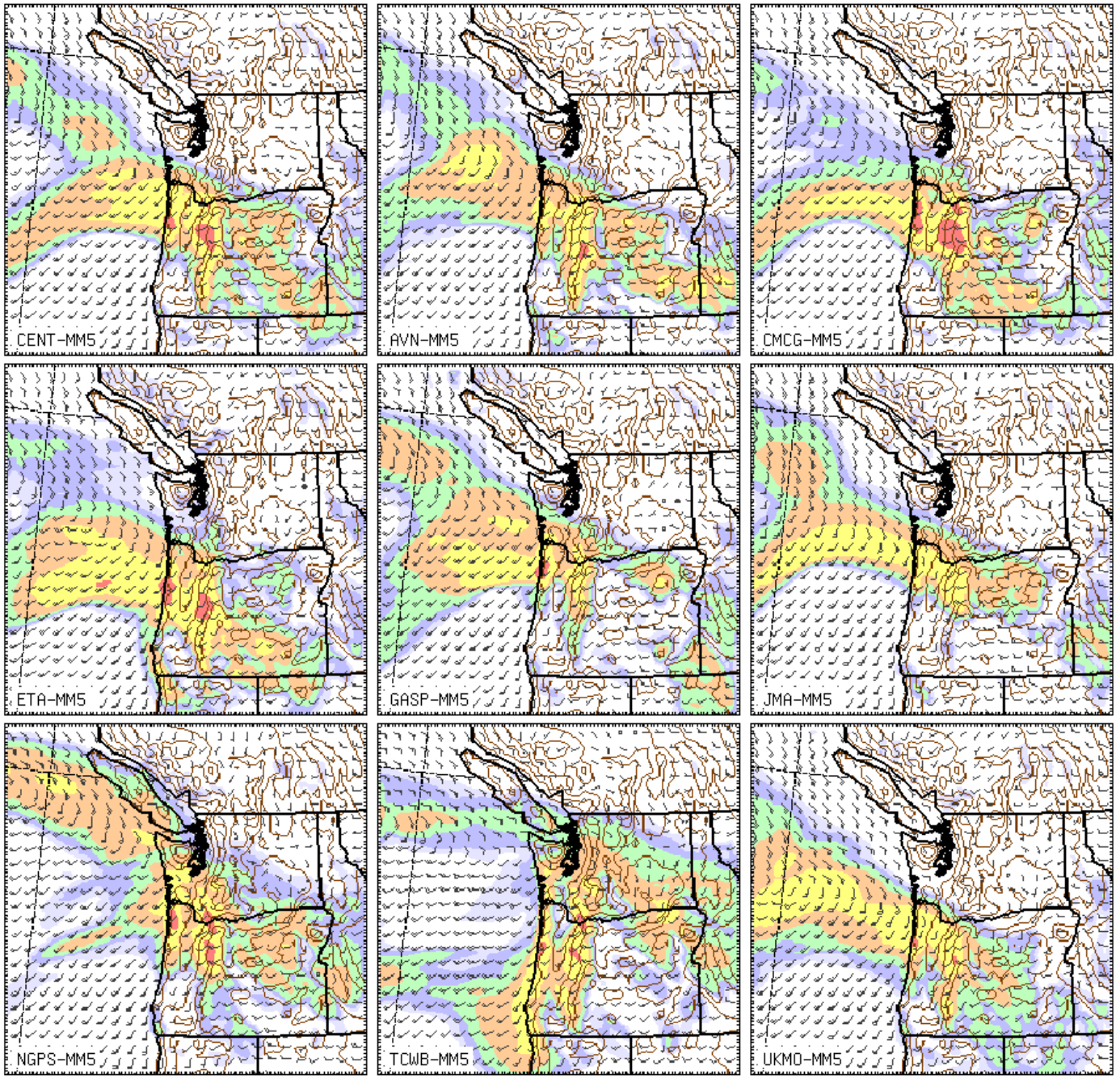
Configurations of the MM5 short-range ensemble grid domains. (a) Outer 151×127 domain with 36-km horizontal grid spacing. (b) Inner 103×100 domain with 12-km horizontal grid spacing.



GFS-MM5 too slow bringing precipitation. Could ensembles have given us some warning?

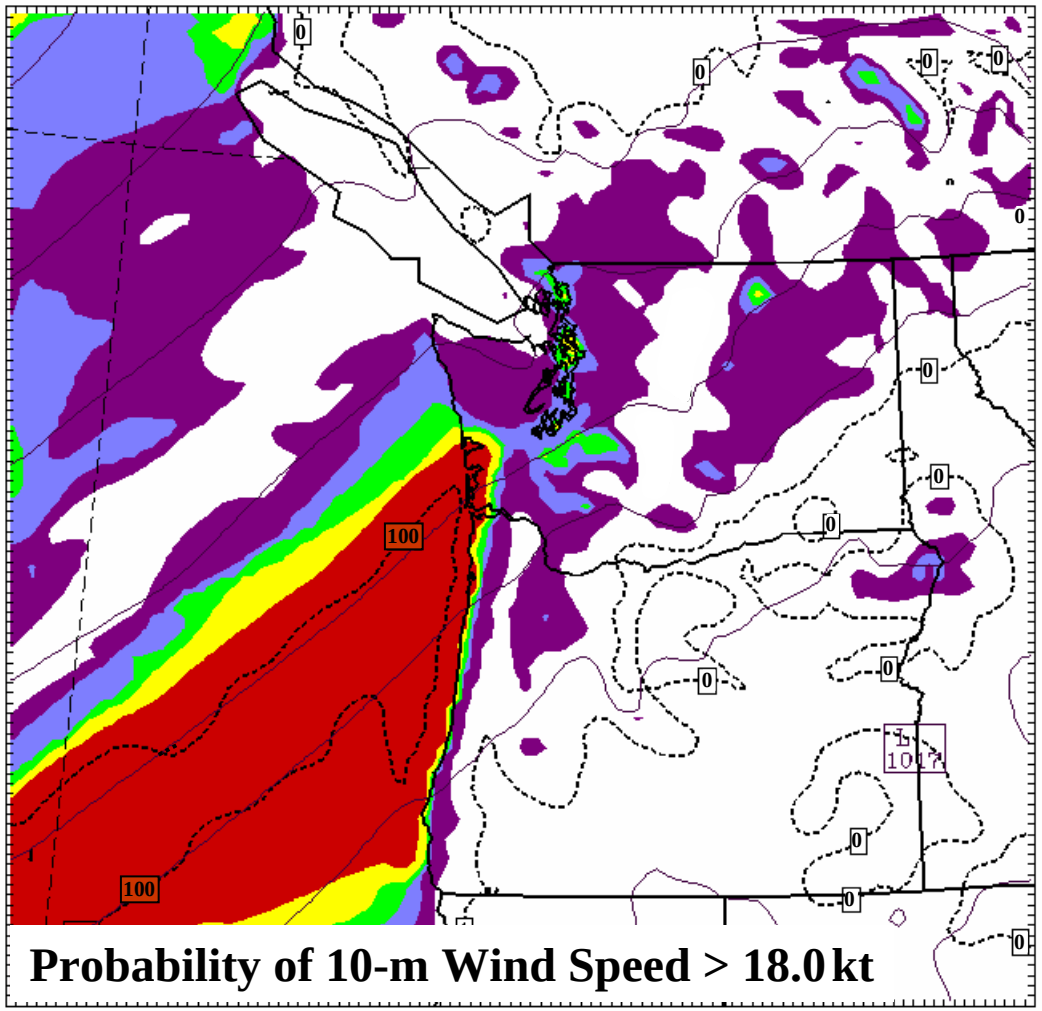


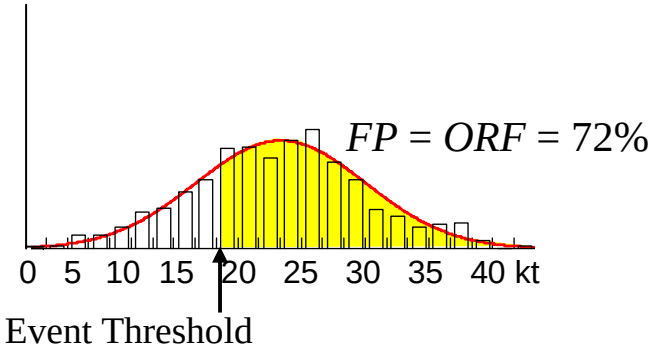
Real-time Deterministic 21-h Forecast


CENT-MM5
AVN-MM5
CMCG-MM5
ETA-MM5
GASP-MM5
JMA-MM5
NGPS-MM5
TCWB-MM5
UKMO-MM5

Mesoscale Forecast Probability

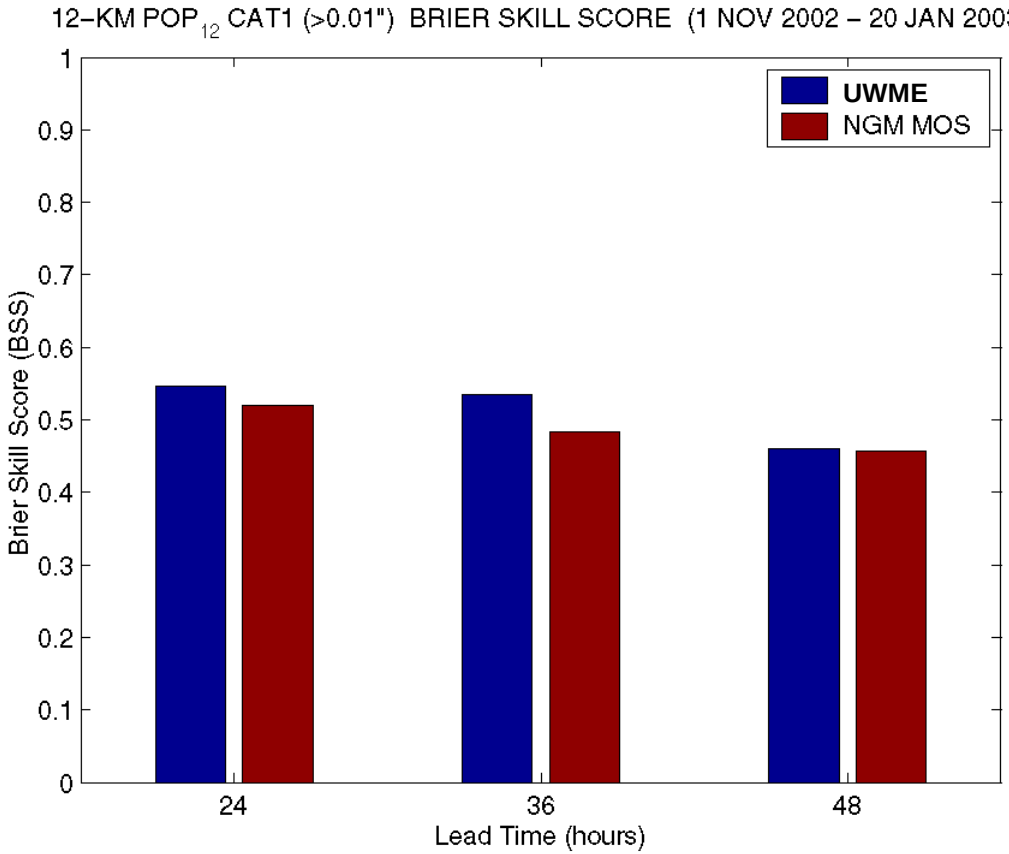
- The primary goal: create skillful, mesoscale forecast probability.
- In a large, ideal EF system, Forecast Probability (*FP*) = Observed Relative Frequency (*ORF*)





- Evaluate FP for a handful of operationally-relevant parameters and thresholds

Mesoscale Forecast Probability



Observation-based verification at MOS locations in 12-km domain.

UWME FP is calculated using Democratic Voting.

- Democratic voting (DV) method provides PoP forecasts as good as NGM MOS.
- Currently Using Uniform Ranks (UR) method. UR is likely better. Calibration would provide further improvements.

UW's Ensemble of Ensembles

	Name	# of Members	EF Type	Initial Conditions	Forecast Model(s)	Forecast Cycle	Domain
Homegrown	ACME	17	SMMA	8 Ind. Analyses, 1 Centroid, 8 Mirrors	"Standard" MM5	00Z	36km, 12km
Homegrown	UWME	8	SMMA	8 Independent Analyses	"Standard" MM5	00Z	36km, 12km
Homegrown	UWME+	8	PMMA	8 Independent Analyses	8 MM5 variations	00Z	36km, 12km
Imported	PME	8	MMMA	8 Independent Analyses	8 operational, large-scale	00Z, 12Z	36km

SMMA: Single Model Multi-Analysis

PMMA: Perturbed-model Multi-Analysis

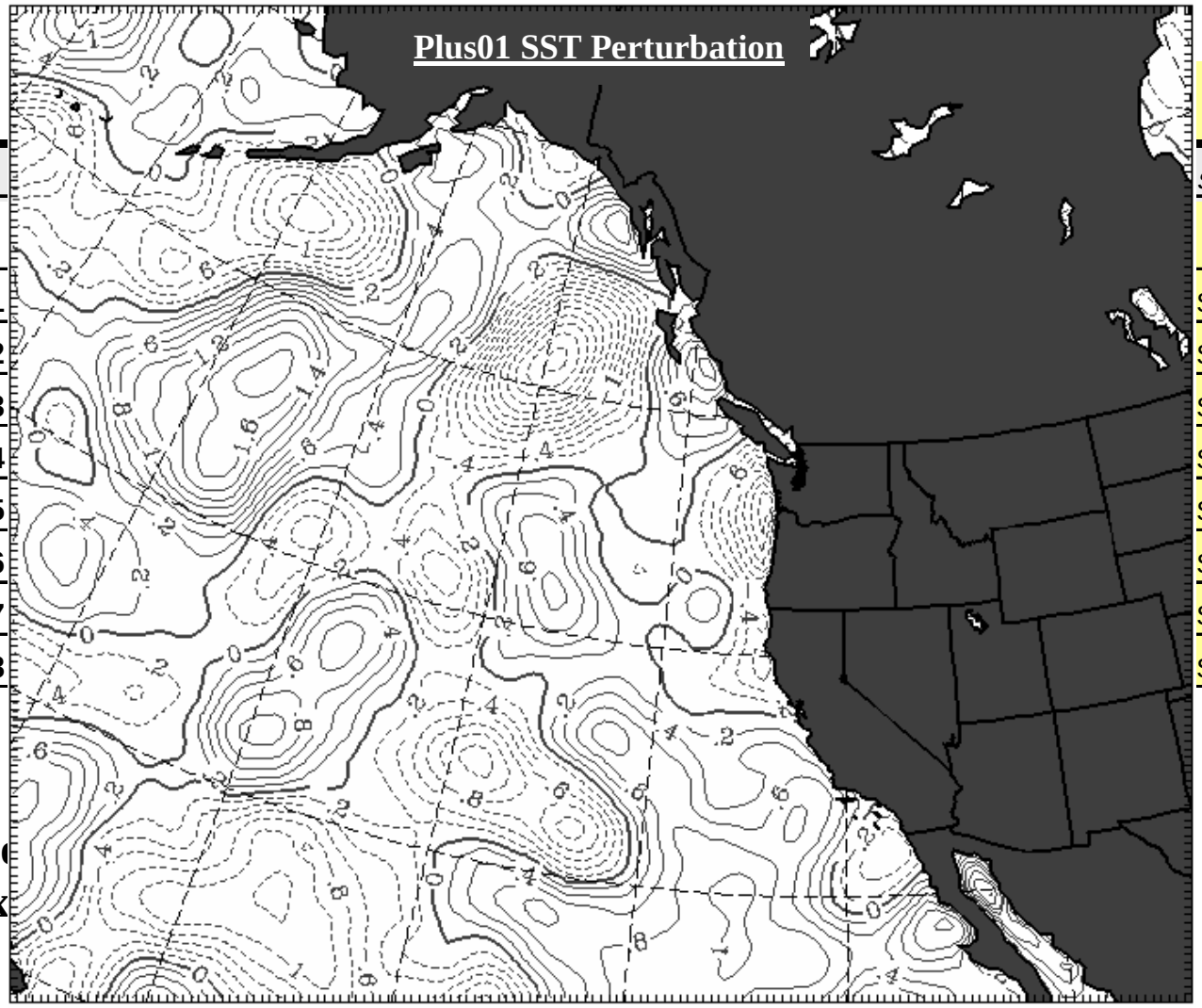
MMMA: Multi-model Multi-Analysis

ACME: Analysis-Centroid Mirroring Ensemble

PME: Poor Man's Ensemble

MM5: 5th Generation PSU/NCAR Mesoscale Modeling System

Design of UWME+



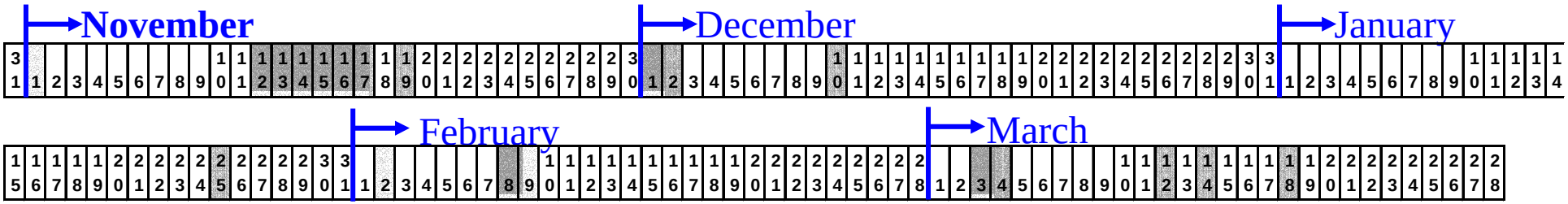
IC	ID#	SST Perturbation	Land Use Table
ACME		*standard*	*standard*
ACMEcore+			
avn	**plus01**	SST_pert01	LANDUSE.plus1
cmcg	**plus02**	SST_pert02	LANDUSE.plus2
eta	**plus03**	SST_pert03	LANDUSE.plus3
gasp	**plus04**	SST_pert04	LANDUSE.plus4
jma	**plus05**	SST_pert05	LANDUSE.plus5
ngps	**plus06**	SST_pert06	LANDUSE.plus6
tcwb	**plus07**	SST_pert07	LANDUSE.plus7
ukmo	**plus08**	SST_pert08	LANDUSE.plus8

1) Albedo
2) Roughness Length
3) Moisture Availability

- **Assume approx**
- **Perturbed surface boundary parameters according to their suspected uncertainty**

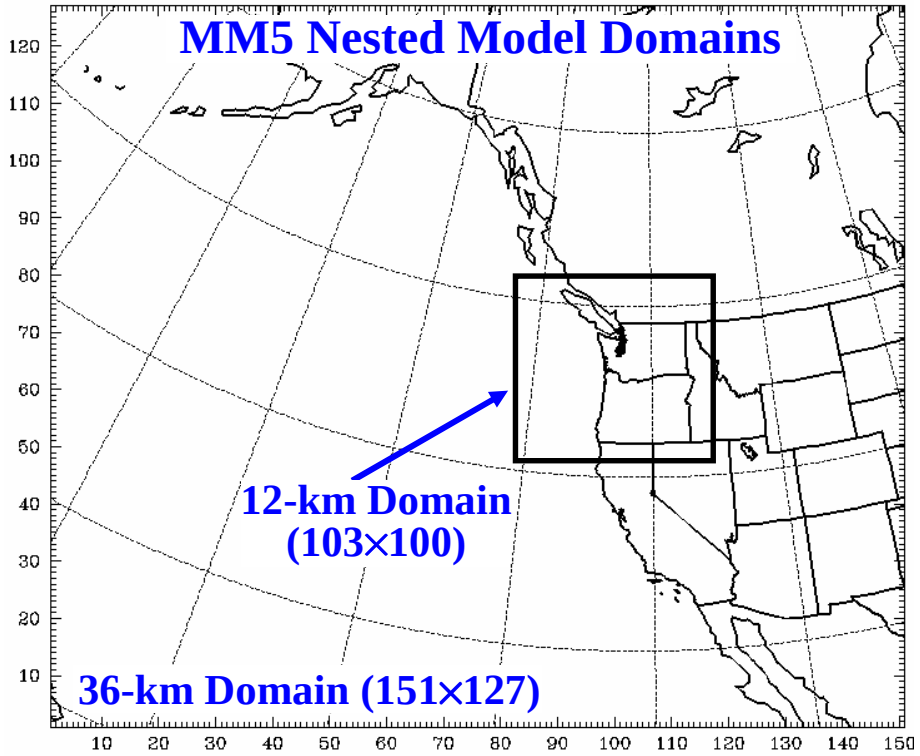
Post-Processing and Verification

Research Dataset



- Total of 129, 48-h forecasts (31 Oct 2002 – 28 Mar 2003) all initialized at 00z
  - Incomplete forecast case days are shaded

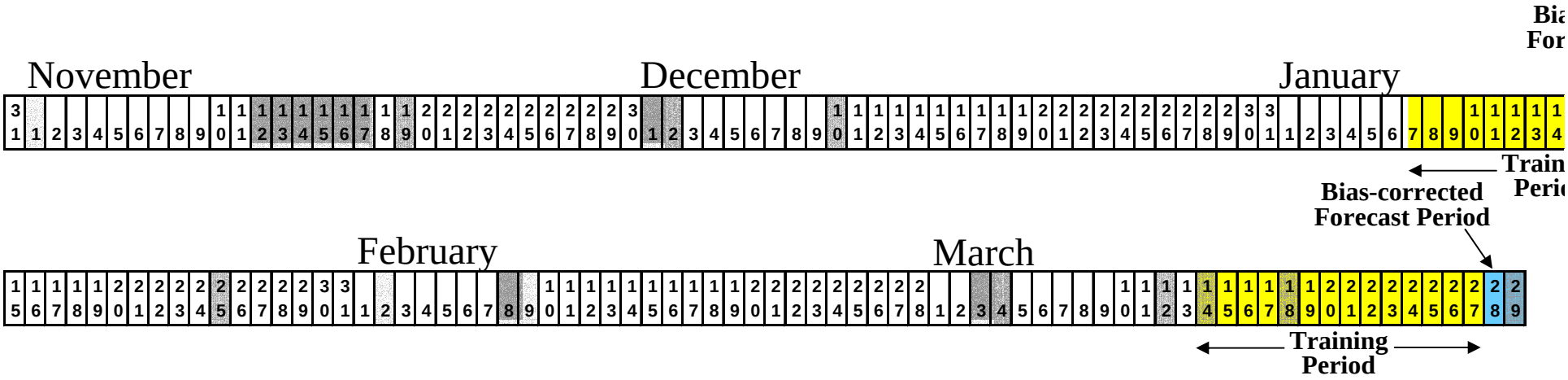
- **Parameters**:
  - 36-km Domain: Mean Sea Level Pressure (***MSLP***), 500mb Geopotential Height ($\boldsymbol{Z_{500}}$)
  - 12-km Domain: Wind Speed at 10 m ($\boldsymbol{WS_{10}}$), Temperature at 2 m ($\boldsymbol{T_2}$)



- **Verification**:
  - 36-km Domain: **centroid analysis** (mean of 8 independent analyses, at 12-h increments)
  - 12-km Domain: **ruc20 analysis** (NCEP 20-km mesoscale analysis, at 3-h increments)

Note: Global PME data was fitted to the 36-km domain

Gridded, Mean Bias Correction



For the current forecast cycle:

1) Calculate bias at every grid point and lead time using previous 2 weeks' forecasts

$$b_{i,j,t} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{f_{i,j,t}}{o_{i,j}}\right)_n$$

N number of forecast cases (**14**)

$f_{i,j,t}$ forecast at grid point ($\boldsymbol{i}, \boldsymbol{j}$) and lead time ($\boldsymbol{t}$)

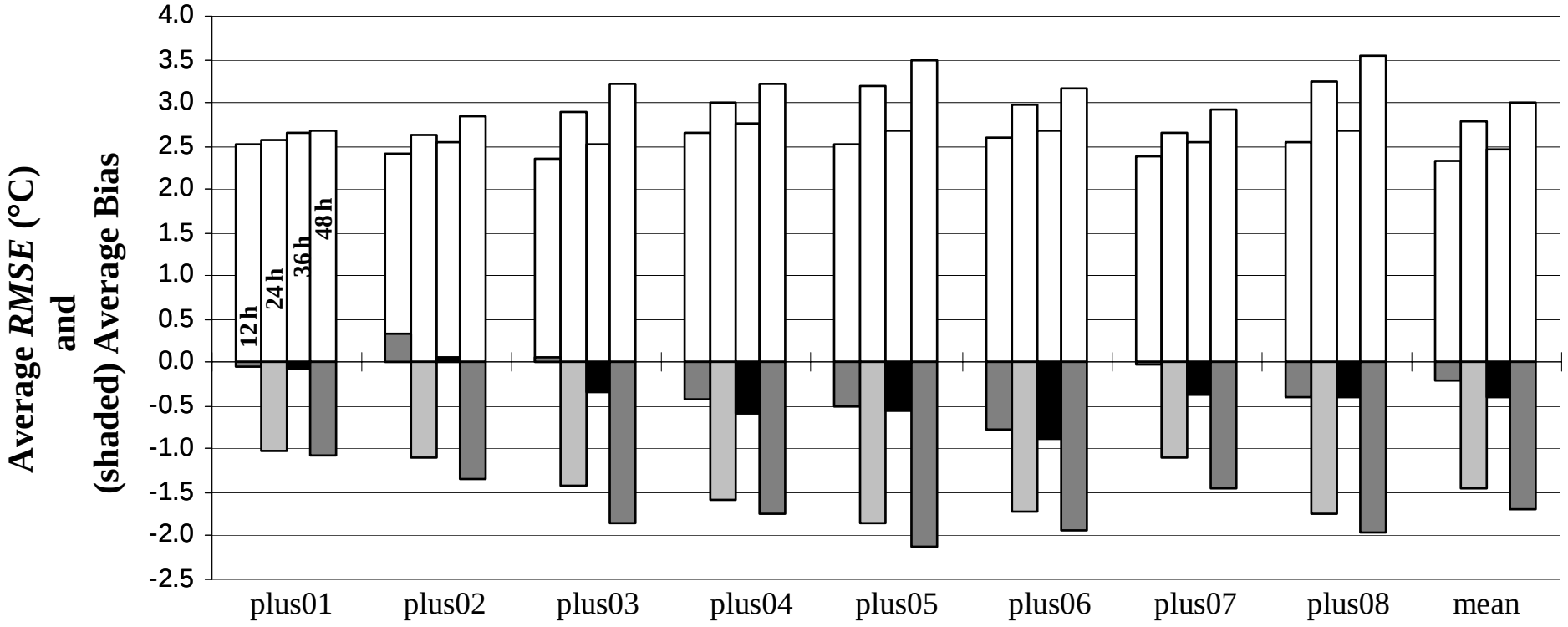
$o_{i,j}$ observation (centroid-analysis or ruc20 verification)

2) Postprocess current forecast to correct for bias:

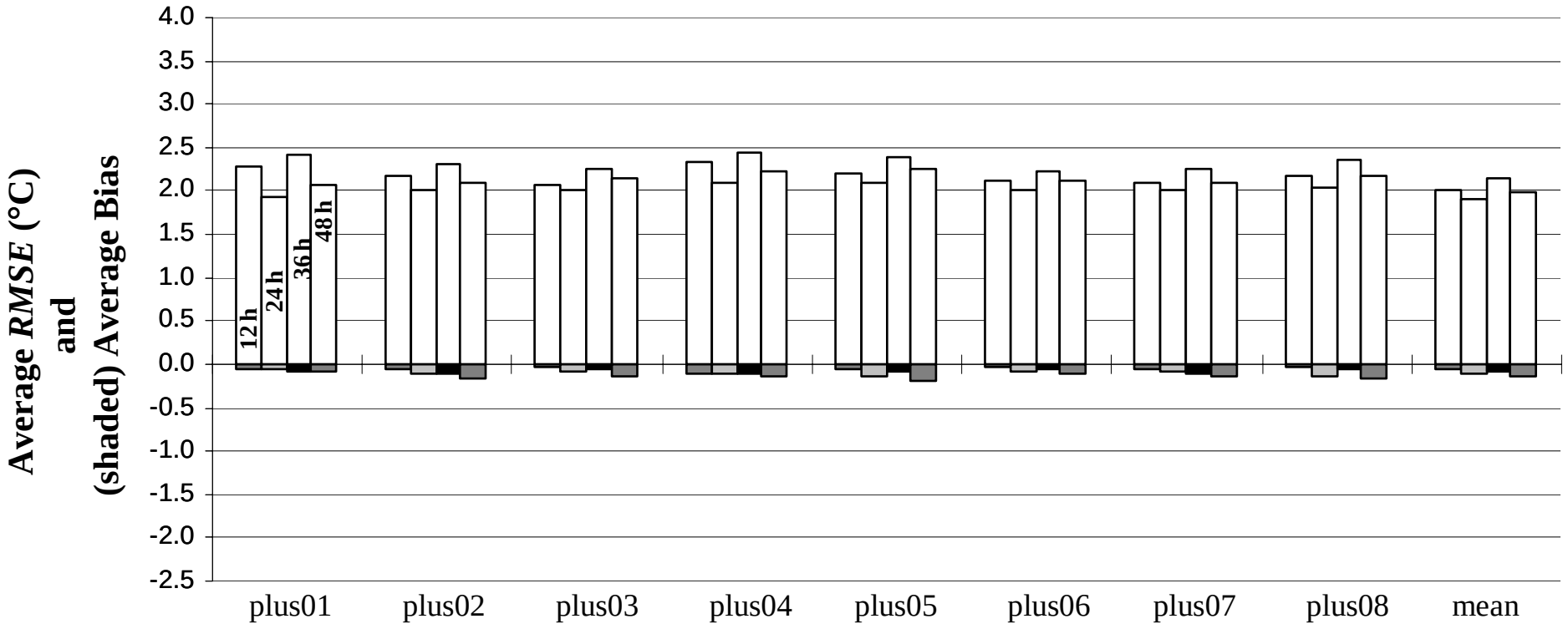
$$f^{*}_{i,j,t} = \frac{f_{i,j,t}}{b_{i,j,t}}$$

$f^{*}_{i,j,t}$ bias-corrected forecast at grid point ($\boldsymbol{i}, \boldsymbol{j}$) and lead time ($\boldsymbol{t}$)

Uncorrected UWME+ T_2

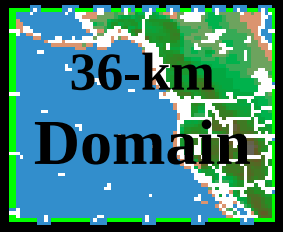


Bias-Corrected UWME+ T_2

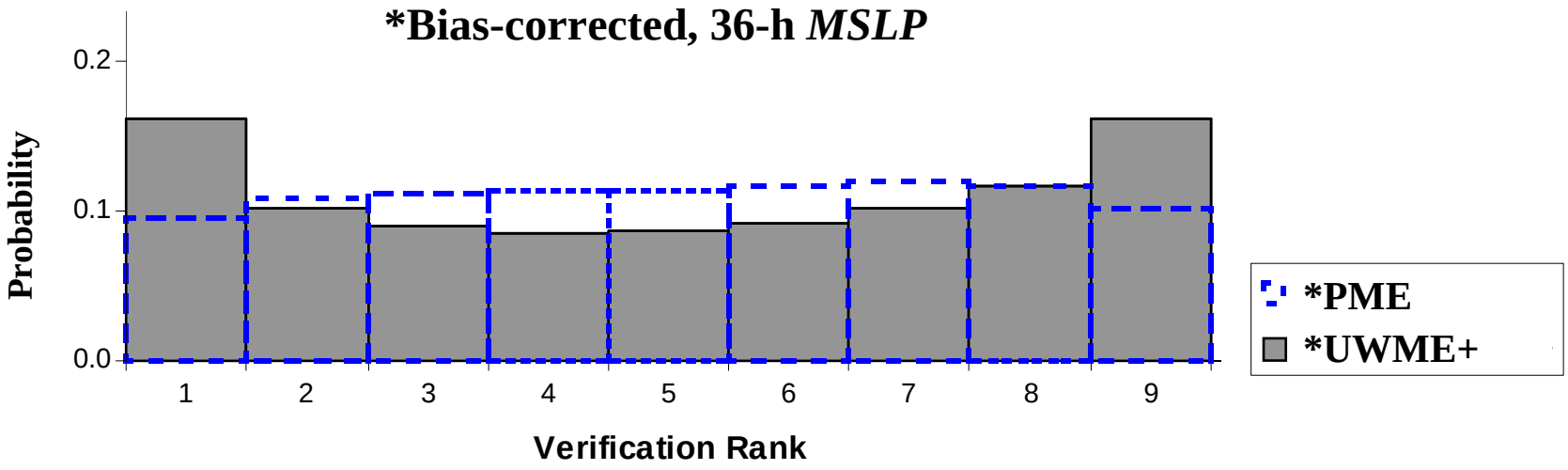


Multimodel
Vs.
Perturbed-Model

PME
Vs.
UWME+



Comparison of VRHs



- ***PME exhibits more dispersion than *UWME+ because**
  - ***PME (a multi-model system) has more model diversity**
  - ***PME is better at capturing growth of synoptic-scale errors**

"Nudging" MM5 outer domain may improve SREF

Verification Rank Histogram

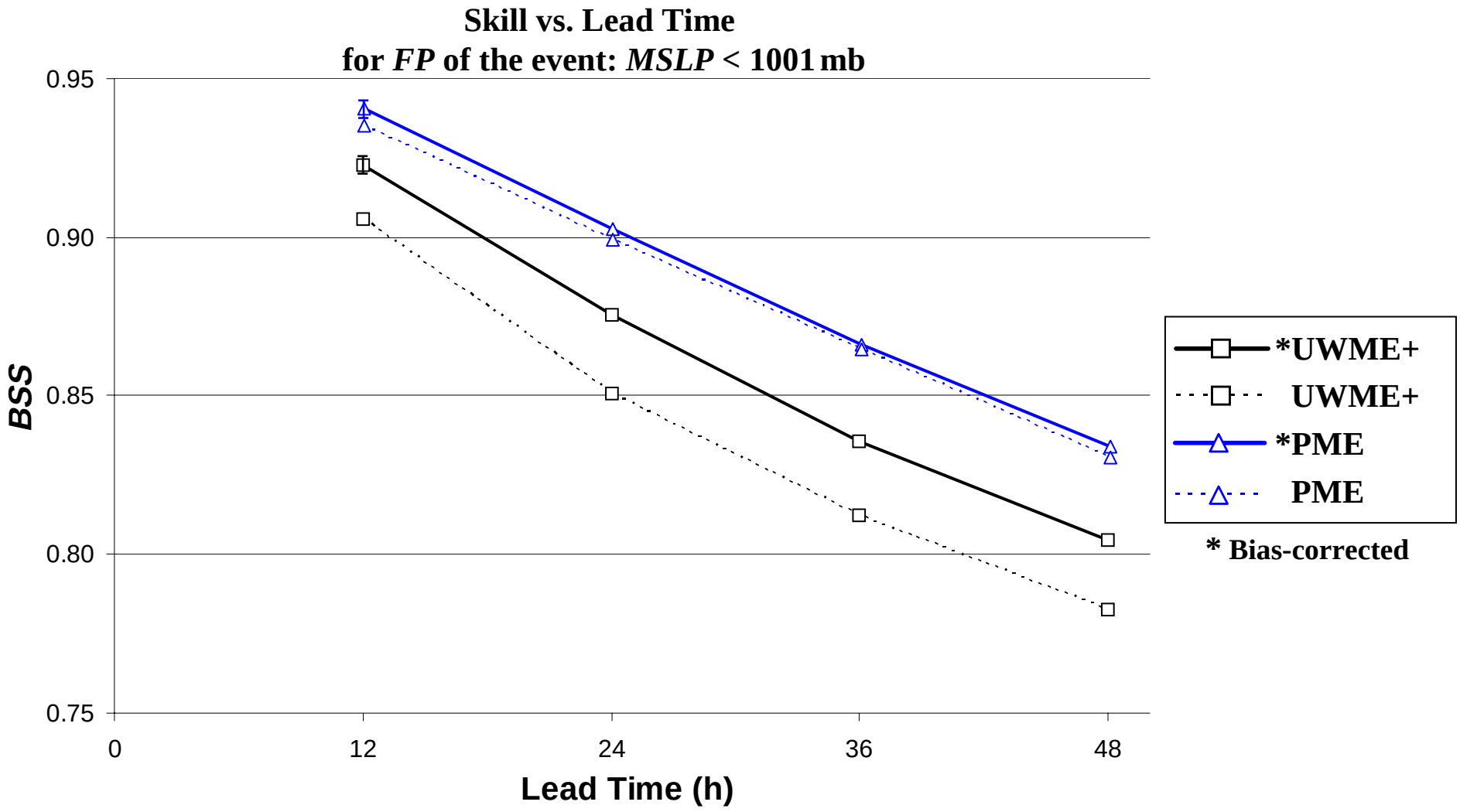
Record of where verification fell (i.e., its rank) among the ordered ensemble members:

Flat  Well calibrated EF (truth's PDF matches EF PDF)

U'd  Under-dispersive EF (truth "gets away" quite often)

Humped  Over-dispersive EF

Comparison of Skill

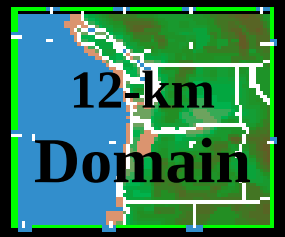


***BSS* = 1, perfect**

***BSS* < 0, worthless**

Value of
Model Diversity
For a Mesoscale SREF

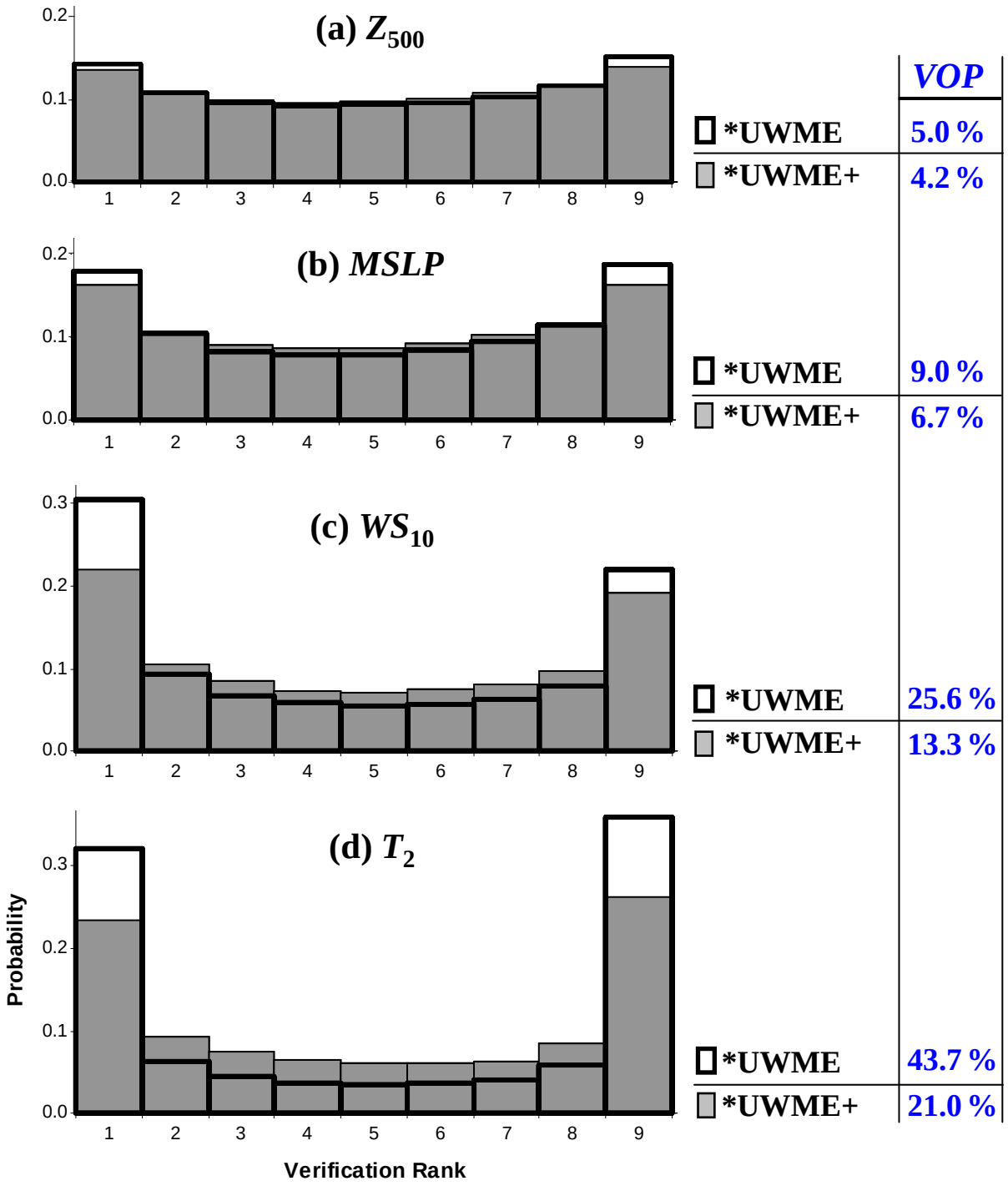
UWME
Vs.
UWME+



Comparison of 36-h VRHs

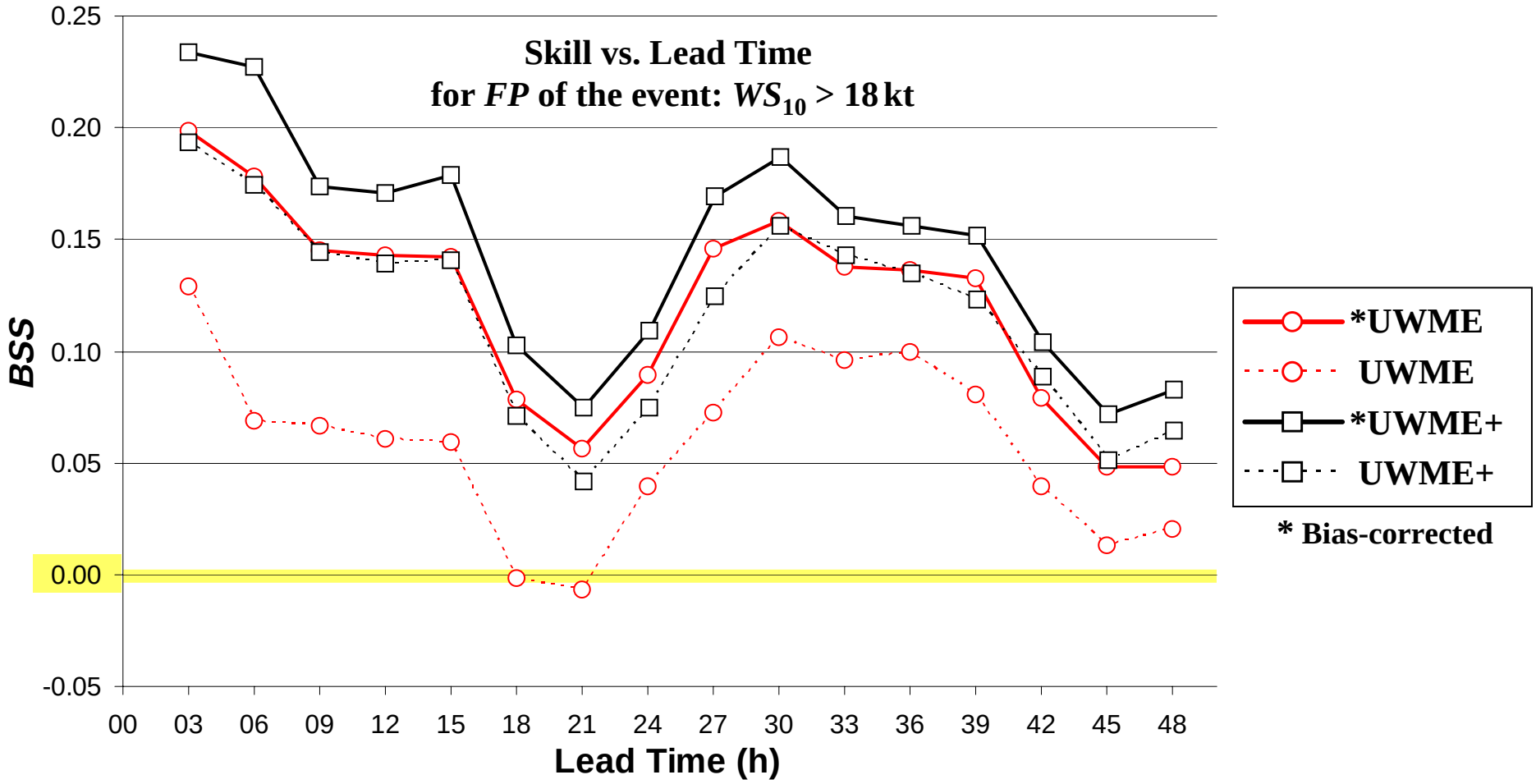
Synoptic Variable
(Errors Depend on Analysis Uncertainty)

Surface/Mesoscale Variable
(Errors Depend on Model Uncertainty)



	VOP
(a) Z_{500} *UWME	5.0 %
(a) Z_{500} *UWME+	4.2 %
(b) *MSLP* *UWME	9.0 %
(b) *MSLP* *UWME+	6.7 %
(c) WS_{10} *UWME	25.6 %
(c) WS_{10} *UWME+	13.3 %
(d) T_2 *UWME	43.7 %
(d) T_2 *UWME+	21.0 %

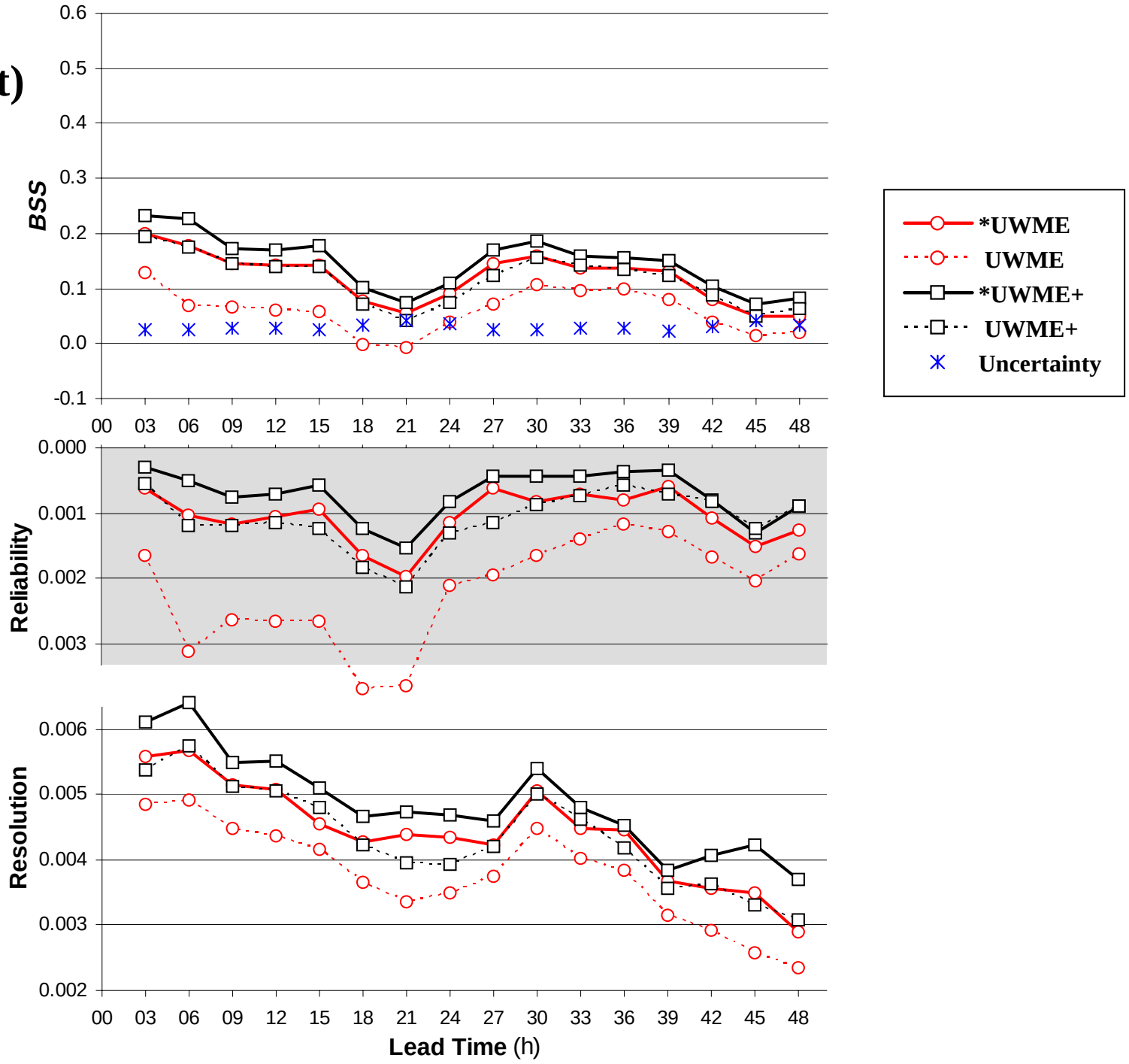
Comparison of Skill



***BSS* = 1, perfect**

***BSS* < 0, worthless**

Skill for $P(WS_{10} > 18\,\text{kt})$



Conclusions

- **Multianalysis Approach for Representing Analysis Uncertainty**
  - **Contributes to a skilled SREF**
  - **Analyses too highly correlated at times—miss key features**
  - **Limits EF size to number of available analyses**
    - **Mirroring produces additional, valid samples of the PDF (i.e., from UWME) but cannot correct deficiencies in original sample**
  - **More rigorous approach would be beneficial to SREF**
    - **UWME is a benchmark for more optimal systems**

- **Bias Correction**
  - **Particularly important for mesoscale SREF where model biases are large**
  - **Significantly improves SREF utility by correctly adjusting the forecast PDF**
  - **Allows for fair and accurate analysis**

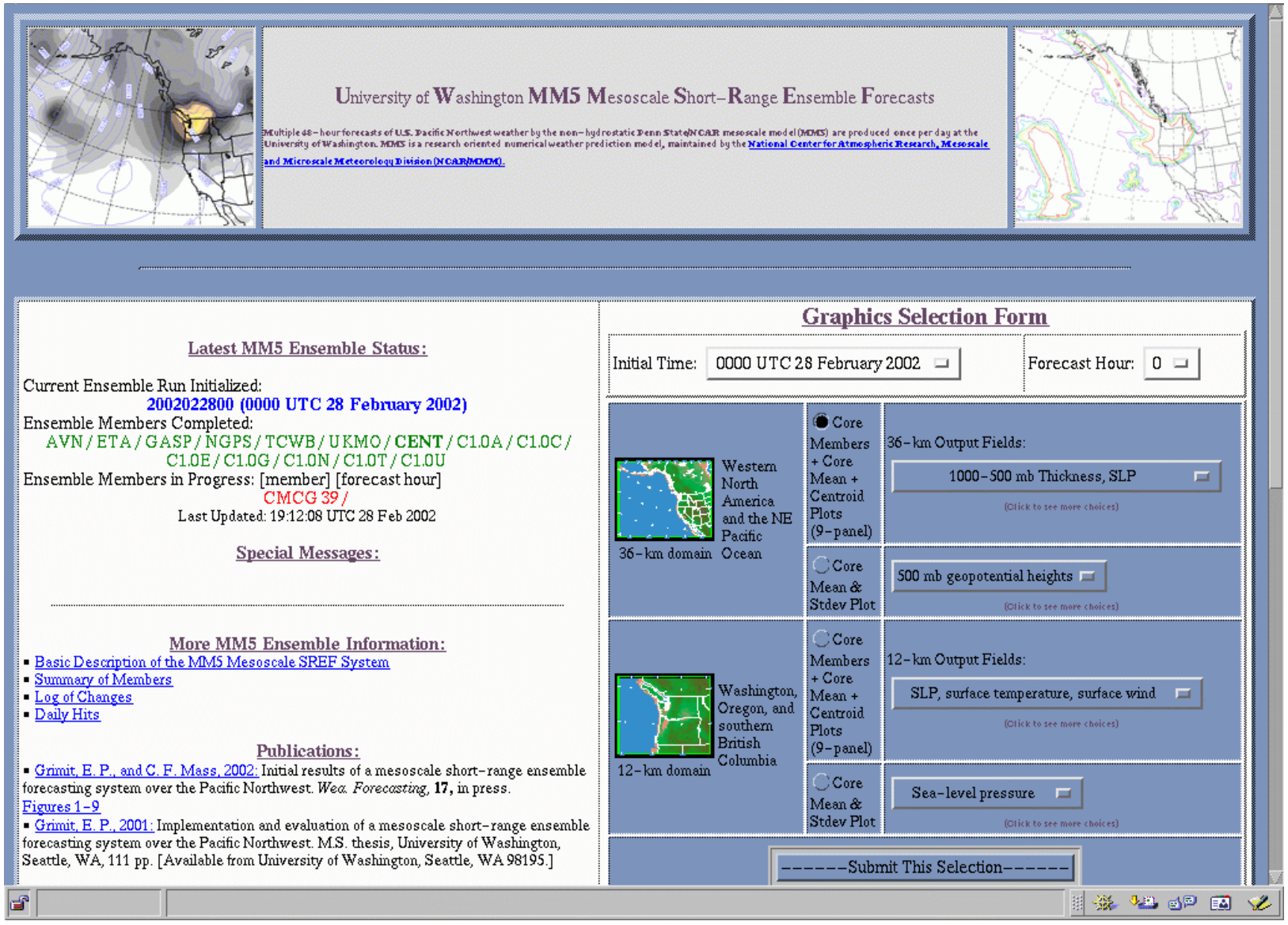
The Future

- UWME is still in its early stages.
  - Several active research projects involving diverse groups
    - NOAA C-STAR (mesoscale EnKF data assimilation, gridded bias removal)
    - DoD MURI (interactive forecast system to handle/visualize forecast uncertainty)
  - Considerable improvement/expansion is planned.
    - Expand to include both cycles (1 May 2004)
    - Implement FDDA/"nudging" on 36-km domain all members (1 May 2004)
    - Implement a number of post-processing approaches: grid-based bias correction, Bayesian model averaging, Ensemble MOS (EMOS)

The Future

- UW can serve as a regional testbed center for mesoscale ensembles.
  - UW can test various mesoscale ensemble and ensemble post-processing approaches for use at NCEP and other modeling centers.
  - UW can test the use of mesoscale ensembles in environmental prediction applications
    - Hydrology/streamflow
    - Air quality
- NOAA-UW relationship is collaborative
  - Experimental UWME products at NWS-Seattle
    - Probability and Mean & Spread data in AWIPS
    - UWME PoP in IFPS
    - Extended-run MM5 for IFPS
  - Would like to compare NCEP SREF with UWME
    - Multianalysis method may be surprisingly difficult to beat
    - Other methods? ETKF?

http://www.atmos.washington.edu/~ens/ensemble.cgi

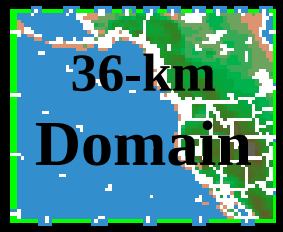


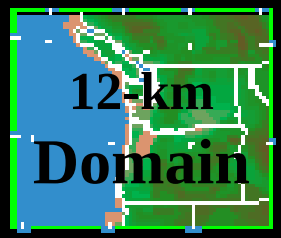
End of Part I

Questions?

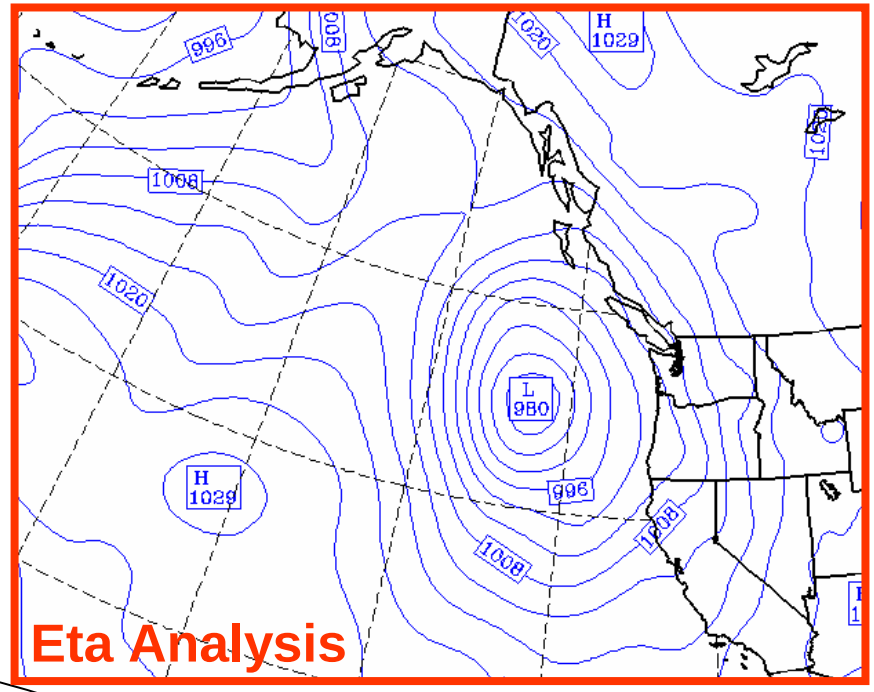
Success and Failure of ACME

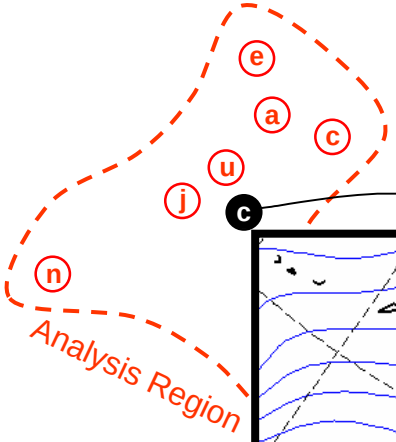
ACME^{core} Vs. ACME

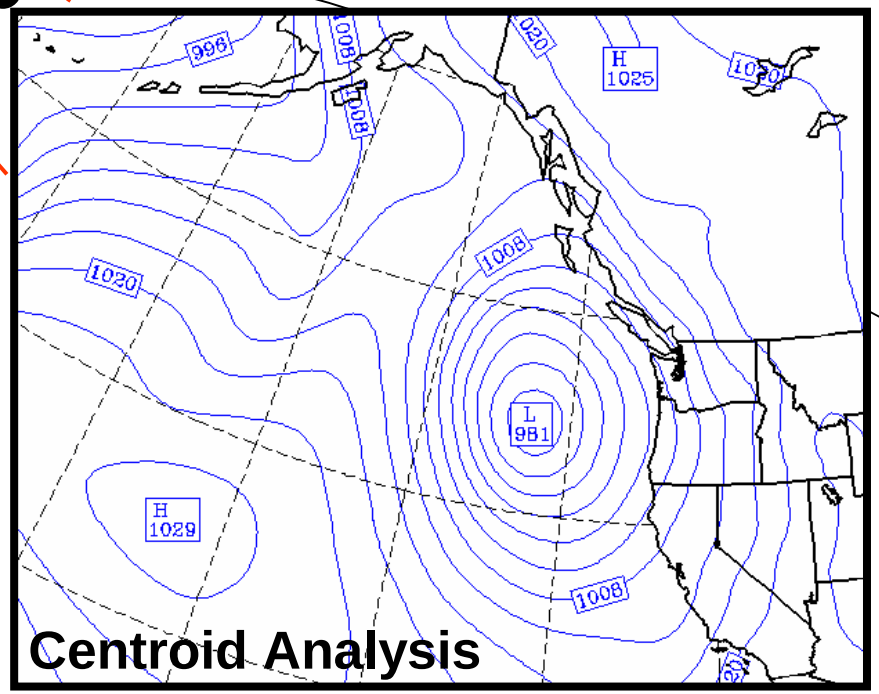


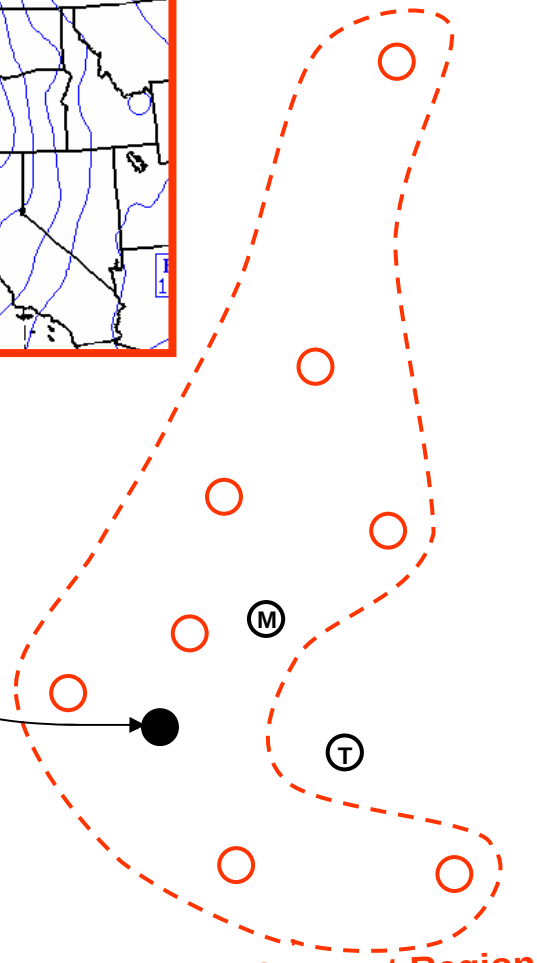


ACME's Centroid Analysis

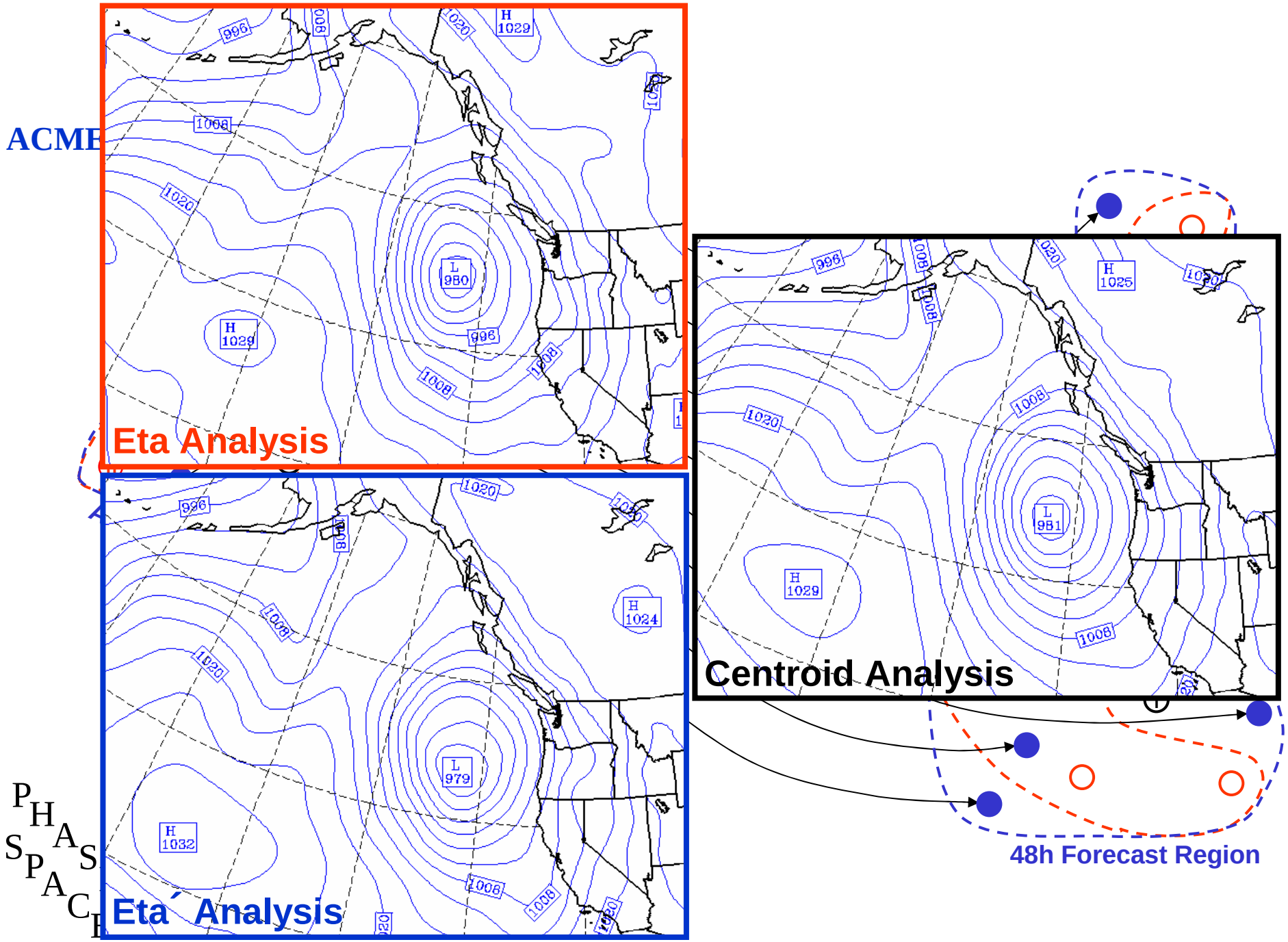




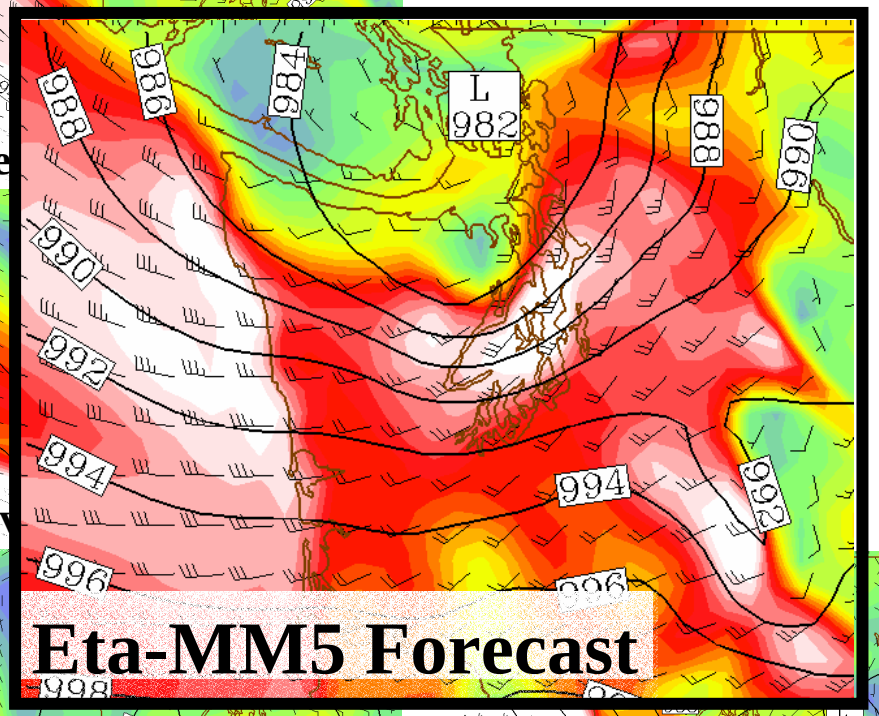




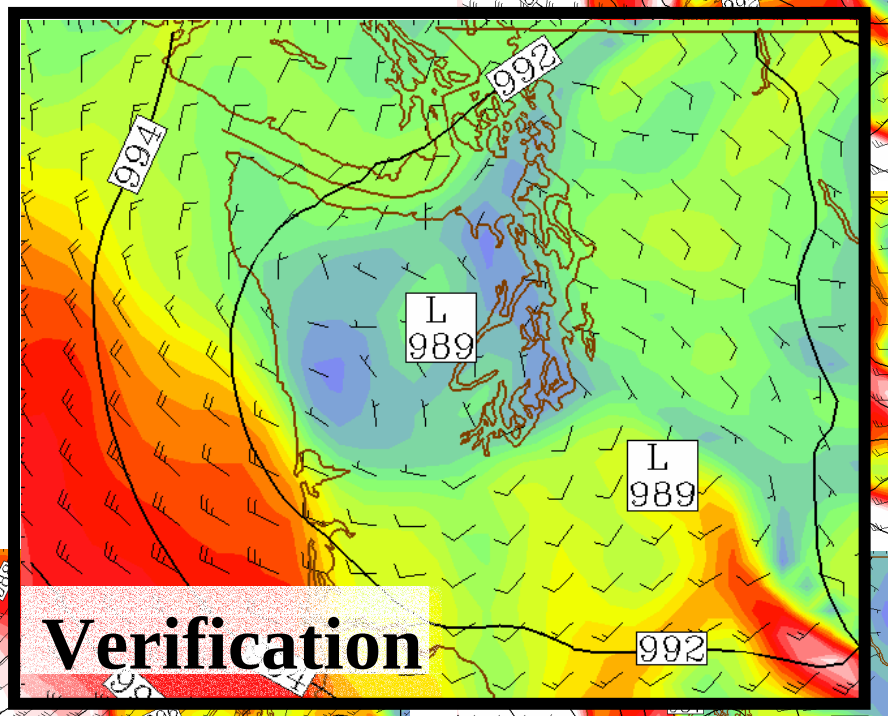
PHASE SPACE


ACME
Eta Analysis
Centroid Analysis
PHASS PACE
Eta´ Analysis
48h Forecast Region

The Wind Storm That Wasn't
(Thanksgiving Day 2001)

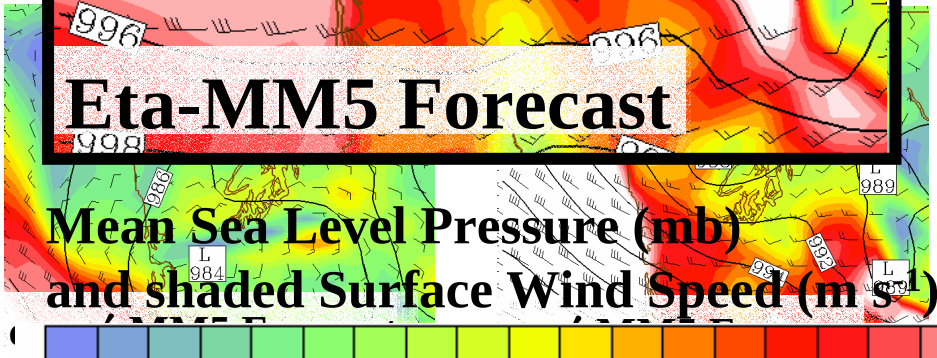


Eta-MM5 Forecast

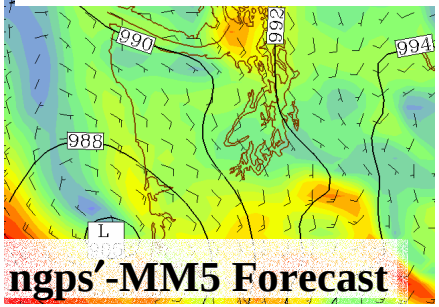


Verification

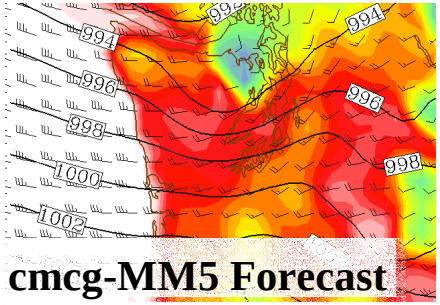
Mean Sea Level Pressure (mb)
and shaded Surface Wind Speed (m s^{-1})



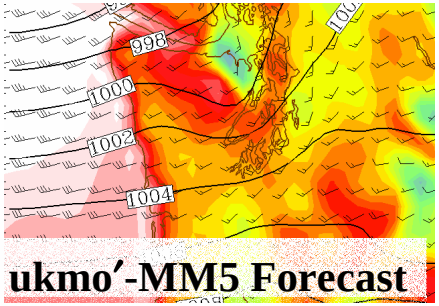
kmo-MM5 Forecast



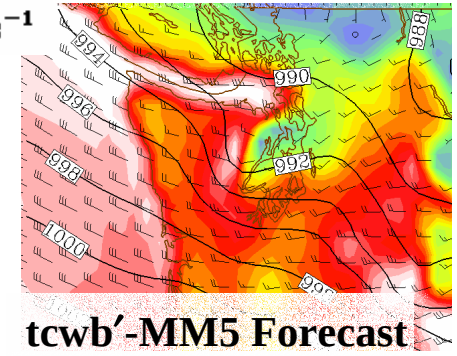
ngps′-MM5 Forecast



cmcg-MM5 Forecast

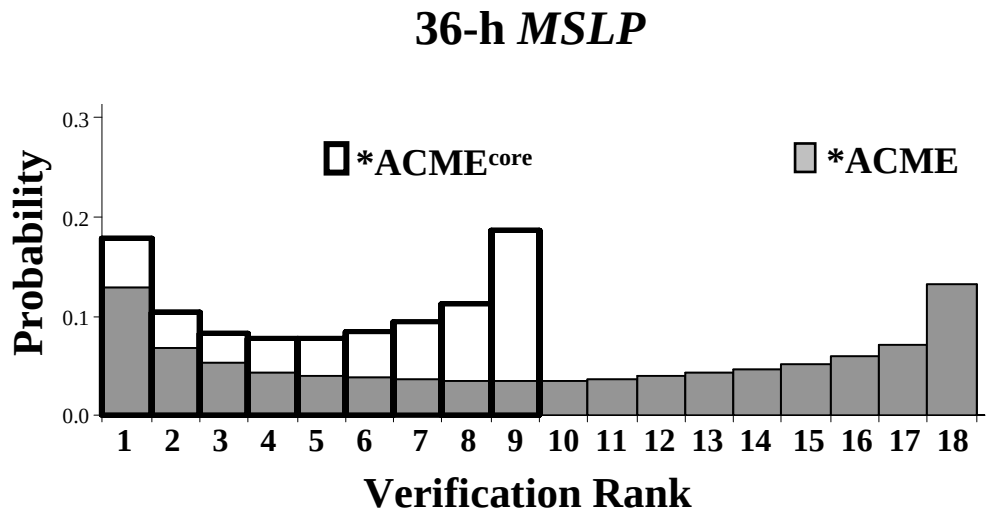


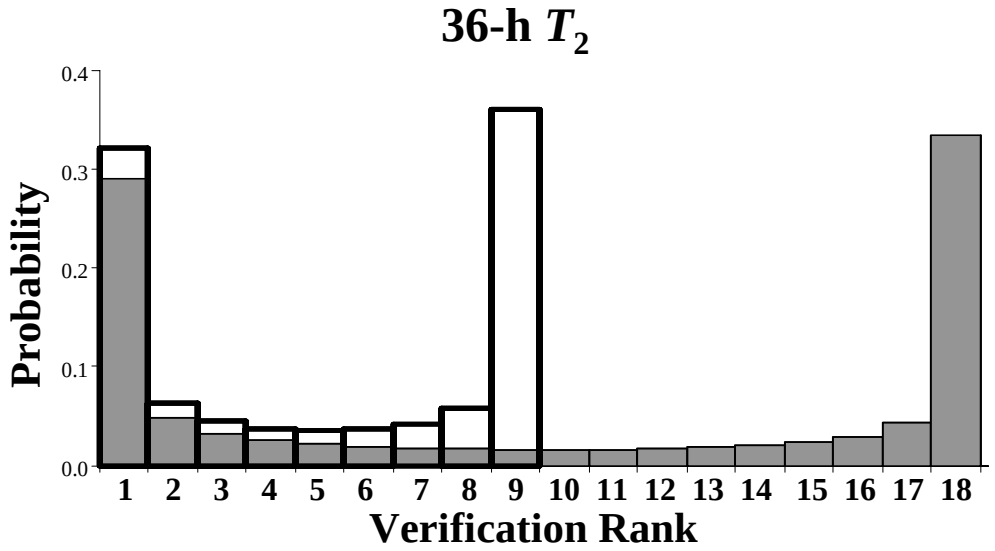
ukmo′-MM5 Forecast



tcwb′-MM5 Forecast

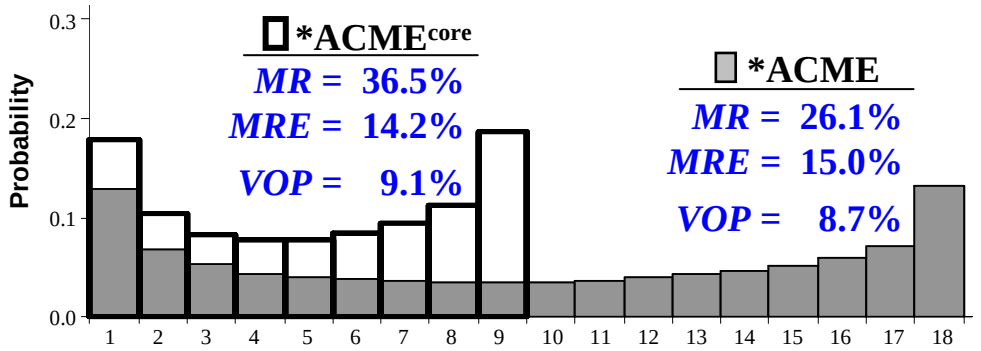
Comparison of Verification Rank Histograms



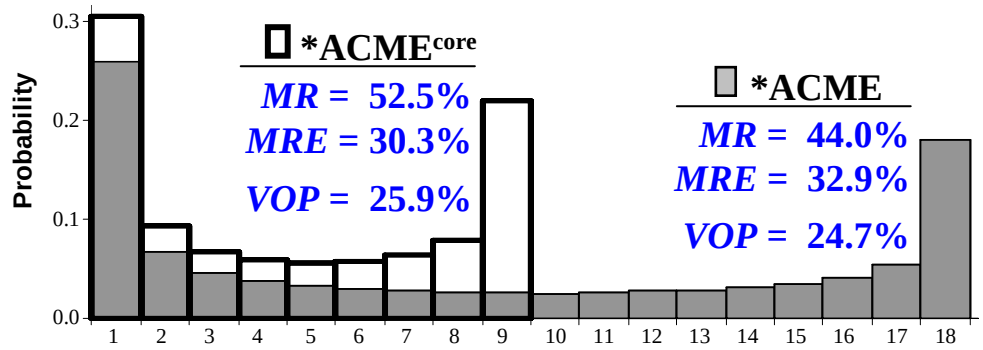


Verification Rank Histograms for…

36-h *MSLP* ------→



36-h WS_{10} ------→



36-h T_2 ------→